NOVEL TRANSCRIPTIONAL REGULATORS AND USES THEREFOR

Priority Claim

The present application claims priority to United States provisional application number 06/056,857, filed August 27, 1997, the entire contents of which are incorporated herein by reference.

Government Support

The work described herein was supported by grants NIGHM22526, NIGHM52409, and NSFMCB9604458 for the United States Government; the United States Government may have certain rights in this invention.

Background of the Invention

Regulated gene expression controls cell differentiation and development. Many medical disorders can be traced to errors in the control of gene expression. Much of biological research is directed toward an understanding of the mechanisms of gene regulation. The holy grail in these studies is the ability to control gene expression.

Very often, gene expression is regulated at the level of transcription. A great deal is known about the machinery involved in gene transcription, and about the factors that regulate it. Most transcriptional regulatory factors are proteins containing at least two functional domains: one that binds to a specific site in DNA and one that interacts, directly or indirectly, with some component of the transcriptional machinery to exert its regulatory effect. These functional domains can typically be separated from one another without losing

their activities, and domains can be swapped from one regulator to another. Also, several known regulators that are naturally found in a particular organism (e.g., yeast) have been found to maintain their activity when introduced into cells of a different organism (e.g., human cells). These observations suggest that at least some transcriptional regulators employ mechanisms that have been conserved in evolution, and that work done in one organism is likely to be generalizable to others.

Despite the large database of accumulated knowledge about the transcriptional machinery and the characteristics of activators, much remains to be learned. For example, there is currently much debate over the mechanism by which regulators interact with and/or alter the activity of the transcriptional machinery. Also, there remains a need for the identification of new and different techniques for controlling gene expression.

Definitions

"Associated" -- The term "associated" is used herein to describe any physical or functional linkage between two moieties. The association may be constitutive or inducible. Moieties may be associated by covalent linkage, hydrogen bonding, van der Waals interactions, hydrophilic or hydrophobic interactions, or any other means that preserves functionality. Where polypeptide moieties are associated by covalent linkage, they form a "fusion protein". Similarly, where nucleic acid molecules are associated by covalent linkage (e.g., by means of a 3'-5' phosphodiester bond, a 2'-5' phosphodiester bond, or some other covalent bond), they form a "fusion molecule". Associations according to the present invention may be direct (i.e., may involve physical contact between or among the relevant moieties) or indirect (e.g., may involve one or more other compounds that mediate(s) the

interaction(s) between or among the moieties). One preferred method for producing an inducible association is to utilize a mediator molecule whose expression or production is itself inducible.

"Isolated" -- A molecule or compound is "isolated" as that term is used herein if is separated from one or more molecules or compounds with which it is associated in nature. Also, any molecule or compound that itself is never produced in nature is by definition "isolated". Any molecule or compound that has been subjected to one or more purification steps is "isolated". Molecules or compounds produced *in vitro* are also "isolated". As used herein, a molecule or compound is "substantially purified" when it is at least about 90% pure.

"Moiety" -- A "moiety", as that term is used herein, is a molecule or a portion of a molecule having a given activity or function. For example, many, if not most, known proteins that regulate transcription have at least two defined functional domains, one that directs DNA binding and one that confers regulation on genes operationally linked with a DNA site to which the protein binds. According to the present invention, the portion or portions of such a protein that are sufficient for DNA binding activity constitute a DNA binding moiety; the portion or portions sufficient for transcriptional regulation constitute a regulatory moiety. Those of ordinary skill in the art will appreciate that a moiety need not be a protein domain, nor a polypeptide of any sort. Any molecule, or collection of molecules, having the designated activity can constitute a moiety according to the present invention.

"Non-naturally occurring" -- The phrase "non-naturally" occurring, when applied to RNA molecules herein, refers to RNA molecules that i) have a nucleotide sequence that is not found in the organism in which they are active; and/or ii) are in a form that is not found in nature. For example, an RNA molecule whose complete nucleotide sequence (i.e., the

sequence of the whole molecule) is not found within any natural RNA molecule endogenous to the cell in which the RNA is activating transcription is non-naturally occurring according to the present definition. Also, any RNA molecule that contains one or more sequence elements that are not naturally found in that cell is a non-naturally occurring RNA molecule. A "sequence element", as used herein, is any portion of an RNA molecule's sequence. All RNA molecules contain one or more sequence elements (since a single "A", "U", "G", or "C" is a sequence element); non-naturally occurring RNAs of the present invention are distinguished by having one or more sequence elements that are not found in naturally-occurring RNAs in the cell in which the non-naturally occurring RNA is active. Alternative non-naturally occurring RNAs are those that, though their nucleotide sequence may be identical to that of an RNA found in nature, are in a form not found in nature. For example, any "isolated" RNA is non-naturally-occurring. Also, any RNA molecule that has been modified so that it includes one or more chemical groups not found in naturally-occurring RNAs of the same sequence, or lacks one or more chemical groups naturally found in RNAs of the same sequence is non-naturally occurring according to the present invention.

"Operationally linked" -- The phrase "operationally linked", as used herein to describe a site in a nucleic acid, means that the site is located in three-dimensional space in a position and orientation that allows a regulator binding to the site to exert its effects on transcription. As is known in the art, transcriptional regulatory sites are commonly linked by covalent association to the promoter whose expression they regulate. That is, regulator sites are typically part of the same DNA molecule as the promoter. In most cases, such sites are located upstream of the promoter, though examples of regulatory sites that are effective when positioned downstream of the promoter, either within or downstream of the gene, are known.

According to the present invention, covalent linkage is not essential for operational linkage between a regulatory site and a promoter. A regulatory site for use in accordance with the present invention may be operationally linked to a promoter, for example, by being provided as a separate DNA molecule or by being embedded within the transcript whose expression the promoter directs.

Summary of the Invention

The present invention provides novel transcriptional regulators comprised of ribonucleic acid (RNA). The RNA regulators of the present invention alter the rate and/or the extent of transcription from a promoter when they are delivered to a site that is operationally linked to that promoter. Those inventive RNA regulators that increase the rate or level of transcription from a given promoter are termed "riboactivators"; those that decrease the rate or level of transcription are termed "riborepressors". Preferred RNA regulators of the present invention are capable of association with a DNA binding moiety characterized by an increased affinity for a particular DNA site as compared with DNA generally. Particularly preferred regulators are characterized by an ability to interact specifically with a TATA-binding protein (TBP) present in the cells in which they are active. Certain preferred riboregulators have a nucleotide sequence that includes a region conforming to the consensus 5'-UGC(G>U>A)GG(U>A>C)(U>ACG)(C>A)(G>A>U)-3' (SEQ ID NO:4).

Description of the Drawings

Figure 1 presents the Wickens/Fields three-hybrid system for identifying proteins that interact with known "bait" RNAs.

Figure 2 presents an inventive two-hybrid system utilized to identify riboactivators.

Figure 3 presents a Northwestern blot demonstrating a specific interaction between riboactivator number 7 and yeast TBP.

Figure 4 presents a Northwestern blot demonstrating that riboactivator number 7 binds to yeast TBP (lane 3), but not to yeast TFIID lacking TBP (lane 4), to mammalian TFIID (lane 2), or to Yeast RNA polymerase II holoenzyme (lane 5).

Description of Certain Preferred Embodiments of the Invention

Riboregulators

Any RNA molecule, or portion of an RNA molecule, that alters transcription from a promoter when recruited to a regulatory site operationally linked with that promoter is a riboregulator of the present invention. Both riboactivators and riborepressors are encompassed within the present invention. Preferred riboregulators affect transcription initiation, but RNAs that affect elongation, reinitiation, termination, and/or pausing are also included within the present invention.

Preferred riboregulators of the present invention are active *in vitro* and/or *in vivo*. Preferably, the riboregulators are active in one or more cell types selected from the group consisting of bacterial cells, yeast cells, mammalian cells, insect cells, plant cells, reptile cells, celenorate cells, and protozoan cells. Particularly preferred riboregulators are active in yeast, mouse, and/or human cells. Preferred riboregulators of the present invention are active at more than one promoter in a particular cell (i.e., affect transcription at various promoters

when recruited to sites operationally linked to those promoters). Particularly preferred riboregulators are active both *in vivo* and *in vitro*.

According to the present invention, preferred riboregulators affect transcription by the same molecular mechanism utilized by one or more naturally-occurring protein factors that regulate transcription. For example, one class of naturally-occurring transcriptional activators, known as "acidic activators", is thought to stimulate transcription by contacting one or more components of the general transcription machinery and recruiting that machinery to the promoter. Although the details of this mechanism are not well understood, a "squelching" assay has been developed to identify activators that fall within this mechanistic category. The term "squelching" refers to the fact that, when a transcriptional activator is expressed at high levels in a cell in which it is active, the effect may be to reduce gene transcription in that cell, sometimes even to the point of killing the cell. Current models explain this observation by hypothesizing that non-DNA-bound molecules of the highly-expressed activator compete with DNA-bound activators for interaction with the same target or targets in the general transcription machinery, so that there is less target available for interaction with DNA-bound activators, the transcription machinery is recruited to promoters less frequently, and transcription is therefore reduced. For purposes of the present invention, any riboactivator that squelches a known acidic activator is classified as one that works by the "acidic activator mechanism". Preferred riboactivators of the present invention squelch, or are squelched by, Gal4 when expressed in yeast and/or in mammalian cells.

In certain preferred embodiments of the invention, the riboregulator includes a stem-loop structure. Particular embodiments of preferred stem-loop containing riboregulators are described in Example 2.

Riboregulators of the present invention may be constitutively active or may alternatively be active only under specified conditions. For example, the techniques and methods described herein may readily be employed to select riboregulators that exert their effects on transcription only in the presence (or absence) of a chemical molecule, a modification enzyme (e.g., a methylase), and/or some component of the growth medium. Riboregulators that are active only in the presence (or absence) of antibiotics may be particularly easy to identify since several RNA-binding antibiotics are known. The techniques of the present invention may also be employed to identify, for example, temperature-sensitive riboregulators. All such conditional riboregulators are encompassed within the scope of the present invention.

DNA binding moieties

As discussed above, inventive riboregulators are RNA molecules that alter the rate or extent of transcription from a promoter when recruited to a regulatory site operationally linked to that promoter. Typically, such recruitment is accomplished by association with a DNA binding moiety. Any chemical compound may be utilized as a DNA binding moiety providing that it has sufficient affinity for DNA to recruit the riboregulator to a regulatory site and also has the ability to form an association with the riboregulator. For example, nucleic acids, polypeptides, intercalation compounds, and/or any chemical compounds that bind to DNA may be utilized as DNA binding moieties. Preferably, the DNA binding moiety is characterized by an increased affinity for a particular regulatory site as compared in DNA generally, so that the DNA binding moiety is said to display sequence-specific DNA binding as that term is understood in the art.

The preferred means by which the riboregulator is associated with the DNA binding moiety depends on the chemical nature of the moiety. For example, where the DNA binding moiety is a DNA or RNA molecule, the riboregulator may be associated with the moiety through covalent linkage (e.g., via a 3'-5' or a 2'-5' phosphodiester bond). In certain preferred embodiments, the riboregulator is produced as a single RNA molecule with the DNA binding moiety. Alteratively, the riboregulator may be associated with a nucleic acid DNA binding moiety through base-pair interactions or other three-dimensional nucleic acid interaction. Where the DNA binding moiety is a polypeptide or a chemical compound, the riboregulator may be covalently linked to the moiety, but is preferably associated by non-covalent interaction.

Applications

Those of ordinary skill in the art will recognize that the riboregulators of the present invention are useful in a wide variety of contexts. For example, they are useful tools in the efforts to understand the mechanism of gene regulation. They are also useful as reagents for identifying and dissecting RNA-RNA and RNA-protein interactions. Furthermore, the inventive riboregulators are useful as agents for controlling gene expression. In one particular embodiment of the present invention, the inventive riboregulators may be used as therapeutic agents to modulate gene expression *in vivo* in order to alleviate or correct a disease state. Certain preferred embodiments of these applications are discussed in more detail below.

Experimental tools

Much effort in the field of transcriptional regulation has been directed toward the identification of the interaction target for the protein regulators that activate and repress transcription *in vivo*. The riboregulators of the present invention provide valuable new tools for this endeavor. Riboregulators that act by the same mechanism as natural transcriptional regulators are particularly useful in this regard. One advantage of riboregulators, as compared with protein factors, for these types of studies is that a variety of RNA-specific techniques are available that can be employed to identify and characterize any interaction partners. For example, riboregulators can be readily cross-linked to any interaction target or targets with which they physically interact. Techniques for cross-linking single-stranded RNAs with nearby proteins are well known in the art. Techniques for cross-linking double-stranded RNAs, which techniques are useful for riboregulators containing stem-loop or other double-stranded secondary structures, have recently been developed (see, for example, Liu et al., in Abstracts of Papers Presented at the 1997 Meeting on Eukaryotic mRNA Processing, Cold Spring Harbor Laboratory, August 20-24, 1997, pg. 82). Other useful RNA-specific techniques include, for example, Northwesterns, gel mobility shift assays, filter binding assays, etc. (see Example 4). Also, complexes containing an RNA molecule can be detected by reverse transcription and amplification.

Riboregulators of the present invention may also be used in novel multiple-hybrid experiments to identify RNA or protein molecules that interact with particular test RNAs. Most available di-hybrid or multiple-hybrid systems rely on protein components and are not readily adaptable to analysis of RNA-RNA interactions. Even RNA-protein interactions are often difficult to study with current systems; only the Wickens system has been developed for that purpose. The inventive demonstration of activating RNAs, however, allows ready

production of multiple-hybrid systems that can analyze RNA-RNA and RNA-protein interactions. For example, the test RNA can be produced as a single hybrid molecule with the riboactivator and then screened against either a protein library fused to a DNA binding domain (a two-hybrid system) or an RNA library fused to an interaction moiety that allows the RNA library to be recruited to DNA (a three-hybrid system). Without a riboregulator, RNA-RNA interactions could only be studied in "four-hybrid" systems, an impractical, if not infeasible arrangement.

Inventive riboregulators may also be employed in multiple-hybrid experiment to identify RNAs that interact with a test protein. In the simplest version of this experiment, the test protein is fused to a DNA binding domain and an RNA library is fused with the riboregulator. Such studies can lead to the development of "RNA linkers" that can be used to bring any two proteins together. That is, once RNAs are identified that interact with first and second proteins, those RNAs are attached to one another (preferably by producing both linked together as a single RNA molecule) to produce an RNA linker. Such RNA linker molecules, although not necessarily riboregulators, are also within the scope of the present invention.

Therapeutic reagents

A large number of medical problems are caused by over- or under-expression of a particular gene, or by mutation of a gene. Many more can be alleviated by increased or decreased gene expression. The inventive riboregulators, which can be specifically targeted to activate or repress a selected gene, are useful as therapeutic agents.

For example, metastatic prostate cancer is currently treated with two drugs, leuprolide, which blocks testosterone production, and flutamide, which acts as an androgen receptor antagonist. These drugs are effective against prostate cancer because prostate cancer cells require hormone stimulation to grow and divide. Unfortunately, these drugs lose their effectiveness after about two years of treatment because resistant cancer cells arise. It is known that at least 80% of these resistant cells express the androgen receptor. Moreover, these cells become resistant to leuprolide and flutamide either by amplifying the androgen receptor gene so that much more receptor is made or by mutating that gene so that the receptor it produces is capable of stimulating cell proliferation even when it is not bound to a hormone ligand. According to the present invention, an improved method of treatment for prostate cancer would involve administration of an agent that represses the androgen receptor gene. Such an agent could be given alone or in combination with leuprolide and/or flutamide. An inventive riborepressor is such an agent. Preferably, the riborepressor is prepared in association with a DNA binding moiety that targets it to a regulatory site that is operatively linked to the androgen receptor gene. Alternatively or additionally, the riboregulator may be associated with an interaction moiety that mediates an association between the riboregulator and a DNA binding entity endogenous to the cells to which the riboregulator is delivered. The riboregulator is packaged into a pharmaceutical formulation according to known techniques and procedures, and is preferably delivered orally.

As but one more example, inventive riboregulators are useful int he treatment of anemia and other disorders related to erythrocyte production. Erythropoietin (Epo), a protein agent that stimulates production of erythrocytes, is perhaps the most successful drug produced through biotechnology. Recent studies have shown that it is possible to make Epo

even more effective by co-administration of an agent, known as interleukin-1, that increases transcription of the erythropoietin receptor (EpoR). Unfortunately, interleukins have proven not to be generally useful as pharmaceutical agents because of their associated toxicities. The present riboactivators provide an alternative agent that could be used to stimulate EpoR gene expression. Because inventive riboactivators can be designed to be highly specific to the particular gene to be activated, they should not have the toxicity problems associated with interleukins, which have broad-ranging activities.

It is appreciated that the structure of the inventive riboregulators may have to be modified in order to ensure their effectiveness as therapeutic agents. It may be necessary, for example, to increase the stability of the riboactivator in order to ensure that it persists long enough *in vivo* to have a significant effect on gene expression. It is known that addition of 2'-0-methyl. or-phosphorothiol groups increases the stability of RNA molecules *in vivo*. Also, circularized versions of the molecules may be more stable. Other modifications might be made in order to increase bioavailability of the riboregulators, and/or to increase the extent to which they are taken up by cells. Those of ordinary skill in the art will appreciate that any of a wide variety of modifications may readily be tested, and the resultant modified RNAs can be assayed as described herein to ensure that they retain their functionality.

Examples

Example 1

A Screen for the Identification of Riboregulators

We have developed a novel two-hybrid transcriptional regulation system for use in the identification of riboregulators. In order to simplify the presentation of our system, we

first describe a prior art system that was prepared by Wickens and Fields to allow the identification of proteins that interact with known RNAs (SenGupta et al., *Proc. Natl. Acad. Sci. USA* 93:8496, 1996).

The Wickens/Fields system, which is depicted in Figure 1, utilizes i) a first hybrid **100** comprising a DNA binding moiety **200** (the lexA DNA binding domain) fused to an RNA binding moiety **300** (the MS2 coat protein); ii) a second hybrid **400** comprising a recruiting RNA **500** (the MS2 RNA) fused to the bait RNA **600**; iii) and a third hybrid **700** comprising a target protein **900** fused to a transcriptional activation domain **900** (the Gal4 activation domain). The DNA binding moiety **200** recognizes a site **1000** positioned upstream of a reporter gene **1100** and recruits the bait RNA **600** to that site because of the interaction between the RNA binding moiety **300** and the recruiting RNA **500**; transcriptional activation results when the bait RNA **600** interacts with the target protein **700**, so that the transcriptional activation domain **800** is brought to the DNA. As with other known two-hybrid (or multi-hybrid) transcriptional activation assay systems, the Wickens/Fields system is designed to screen protein libraries (specifically, libraries of target proteins) to identify those with a desired interaction capability (i.e., the ability to interact with the known bait RNA).

We have modified the Wickens/Fields system as shown in Figure 2. Our system employs only two hybrids, one **100** comprising a DNA binding moiety **200** fused to an RNA binding moiety **300**, and one **400** comprising a recruiting RNA **500** linked to a bait RNA **600** (the riboregulator). In this system, the bait RNA **600** is recruited to the DNA because of the interaction between the RNA binding moiety **300**, which itself is bound to DNA because of its linkage to the DNA binding moiety **200**, and the recruiting RNA **500**; transcriptional

activation results whenever the bait RNA **600** is a riboactivator; transcriptional repression results when the bait RNA **600** is a riborepressor.

The inventive system depicted in Figure 2 differs from the Wickens/Fields system in a least two major ways: i) it does not rely on a third hybrid, containing a transcriptional activation domain, to mediate transcriptional regulatory effects but rather detects transcriptional regulation directed by the bait RNA itself; and ii) the bait RNA is a library of RNA molecule from which active RNAs are identified, rather than a known RNA that is used to identify active interacting proteins.

In the particular embodiment of the inventive system that is depicted in Figure 2, the DNA binding moiety **200** consists of a lexA DNA binding domain; the RNA binding moiety **300** is the MS2 protein; and the recruiting RNA **500** is the MS2 RNA. In light of the above discussion, those of ordinary skill int he art will appreciate that each of these system components could readily be substituted with a different molecule (or group of molecules) that performs substantially the same function. For example, any molecule or compound that is capable of both (i) specific recognition of a DNA sequence and ii) interaction with an RNA binding moiety, when utilized under the conditions of the assay, is useful as a DNA binding moiety in accordance with the present invention.

Also, any of a variety of molecules or compounds may be utilized as RNA binding moieties and recruiting RNAs in place of the MS2 protein and MS2 RNA depicted in Figure 2. Useful RNA binding moieties must interact specifically with both the DNA binding moiety and the recruiting RNA under the conditions of the assay; useful recruiting RNAs must interact specifically with the RNA binding moiety and the bait RNA.

Any of a variety of different genes may be utilized as reporter genes, so long as their gene products are detectable. In screens for riboactivators, it is generally preferred that the RNAs be tested for their ability to activate at least one reporter gene that is an essential gene under the conditions of the assay (so that the assay is a selection, rather than a screen). In screens for riborepressors, it is preferred that the RNAs be tested for their ability to repress a gene whose expression is toxic under the conditions of the assay. In such assays, it may also be desirable to arrange that the cell's sensitivity to the toxic gene be inducible, in order to avoid the problem of the cell being killed by expression of the gene before the potential riborepressors have had the opportunity to exert their effects.

When using the screen depicted in Figure 2, an RNA is identified as a riboregulator if it exerts any reproducible effect (increase or decrease beyond that expected error) on transcription. Preferred riboactivators increase or decrease gene expression by at least 2-fold, preferably at least 5-fold, more preferably at least 10-20 fold, yet more preferably at least 100 fold, at most preferably at least 1000 fold.

Example 2

Identification and Characterization of Riboactivators from an "R10" Library

Materials and Methods

STRAINS AND MEDIA: The *L40*-coast yeast strain (a derivative of the *ura3 L40* strain containing the lexA-MS2 fusion protein integrated into the genome; *MATa, ura3-52, leu2-3, -112, his3Δ200, trp1Δ1, ade2, LYS2::(lexA-op)-HIS3, ura3::(lexA-op)-LacZ, LexA-MS2 coat protein (TRP1))* was generously provided by Marvin Wickens. Rich (YPD) and synthetic (S)

media were prepared as described by Rose et al. (*Methods in Yeast Genetics*, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1990).

CONSTRUCTION OF THE R10 LIBRARY: An oligonucleotide, called "Oligo2" that included a 15nt stem-loop and a randomized 10-mer loop was synthesized. Oligo 2 had the sequence 5'-

CTCTGGGAGCTGCGATTGGCAGAATTCCGGCTAGAACTAGTGGATCCCCCGGGCG

AGGCTTATCCNNNNNNNNNNGGATGTGCTGACCCCGGGCAGCTTGCATGCCTGC

AGGTCGACTCTAGAAAACATGAGGATCACCC (SEQ ID NO:1). Two other oligonucleotides, "Oligo 1" and "Oligo 3" were also prepared. The sequences of these oligonucleotides were selected so that they could be used to amplify Oligo 2 to produce a template library that, when transcribed, would produce the "R10" library of RNA molecules. Oligo 1 and Oligo 3 had the following sequences: 5'-CTCTGGGAGCTGCGATTGGC (SEQ ID NO:2) and 5'-GGGTGATCCTCATGTTTTCT (SEQ ID NO:3), respectively.

Oligo 2 was amplified in ten simultaneous polymerase chain reactions utilizing 72 uL water, 109 uL 10 X PCR buffer (Promega), 4 uL magnesium chloride (Promega, 2 uL nucleotide mix (10 mM each), 3 uL Oligo 1 (500 ng/uL), 3 uL Oligo 3 (500 ng/uL), 5 uL Oligo 2 (50 ng/uL), and 1 uL Taq DNA polymerase each. Following one cycle of 5 minutes at 94 °C, 2 minutes at 50 °C, and 3 minutes at 72 °C, thirty cycles of 1 minute at 94 °C, 2 minutes at 50 °C, and 3 minutes at 72 °C were performed. A final 10 minute extension reaction at 72 °C was then performed. Products were purified over a G25 spin column and were ethanol precipitated. Products were then resuspended in sterile 1 X TE, to a final volume (from all 10 reactions) of 100 uL.

YEAST TRANSFORMATION. Yeast transformation was performed by the lithium acetate method (Rose et al., *Methods in Yeast Genetics,* Cold Spring Harbor Press, Cold Spring Harbor, NY, 1990), exploiting the gap repair phenomenon described by Rothstein (*Methods Enzymol*, 194:281, 1991). 1 μg, 5 μg, 10 μg, or 20 μg of R10 library PCR product were introduced into yeast cells in combination with SmaI-cut pIII/MS2 vector. Transformants were selected on S media lacking the appropriate amino acids.

X-GAL PLATE SCREEN: Transformants were grown on white nitrocellulose filters placed on an appropriate selective medium. Subsequently, the filters were transferred to a synthetic medium containing 2% glycerol and 50 μg/ml x-gal. Pale blue to dark blue colonies were identified, the plasmids they contained were isolated according the known procedures (Ward, *Nuc. Acids Res.*, 1990), and the isolated plasmids were retransformed into the same strain and were retested for transcriptional activation activity.

LIQUID β-GALACTOSIDASE ASSAYS: Transformants that turned blue on the X-gal plate screen were grown to an OD_{600} of between 1.0 and 2.0 in 5 mls of selective media containing 2% glucose. β-galactosidase activity was assayed as described (Rose et al., *Methods in Yeast Genetics*, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1990); units of activity were calculated as nmol/min/mg protein. Each assay was performed at least in duplicate on several occassions.

CONSTRUCTION OF STRAIN FOR GAL4-MS2 EXPERIMENTS: A blaster plasmid (Alani et al., *Genetics* 116:541, 1987) was utilized to disrupt the *TRP* gene of the GGY1 strain (Gill et al., *Nature* 334:721, 1988) by integration of the *URA3* gene. *URA3* gene was deleted by recoving yeast colonies that could grow on plates including 5-FOA.

PRODUCTION OF KEENE RNA-MS2 FUSION: We ordered two oligonucleotides, Keene-1 (5'-CCGGGCGAGGCTTATCCTGGTGGAGCAGGATGTGCTGACC; SEQ ID NO:18) and Keene-2 (5'-CCGGGGTCAGCACATCCTGCTCCACCAGGATAAGCCTCGC; SEQ ID NO:19). These oligos were annealed to one another and were cloned into the *Xma*I site of pMS2-2 (SenGupta et al., *Proc. Natl. Acad. Sci. USA* 93:8496, 1996). A number of clones were sequenced; one clone that had insert in the proper orientation was tested for transcriptional activation capability in yeast.

PRODUCTION OF TAR-MS2 FUSION: We ordered two oligonucleotides, Tar-1 (5'-CCGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGAGAAC; SEQ ID NO:20) and Tar-2 (5'-CCGGGTTCTCTAGTTAGCCAGAGAGCTCCCAGGCTCAGATCTGGTCTAACCAGAGAGAC; SEQ ID NO:21). These oligos were annealed to one another and were cloned into the *Xma*I site of pMS2-2 (SenGupta et al., *Proc. Natl. Acad. Sci. USA* 93:8496, 1996). A number of clones were sequenced; one clone that had insert in the proper orientation was tested for transcriptional activation capability in yeast.

PRODUCTION OF NUMBER 5-TAR STEM-MS2 FUSION: We ordered two oligonucleotides, 5/Tar-1 (5'-CCGGGTCTCTCTGGTTAGACCAGATCTGAGCCGGGATGCTCTCTGGCTAACTAGAGAAC; SEQ ID NO:22) and 5/Tar-2 (5'-CCGGGTTCTCTAGTTAGCCAGAGAGCATCCCGGCTCAGATCTGGTCTAACCAGAGAGAC; SEQ ID NO:23). These oligos were annealed to one another and were cloned into the *Xma*I site of pMS2-2 (SenGupta et al., *Proc. Natl. Acad. Sci. USA* 93:8496, 1996). A

number of clones were sequenced; one clone that had insert in the proper orientation was tested for transcriptional activation capability in yeast.

FUSION OF NUMBER 7 LOOP TO 6-BASE PAIR *SMA*I STEM: We ordered two oligonucleotides, 7/*Sma*-1 (5'-CCGGGTGCTGGATCAC; SEQ ID NO:24) and 7/*Sma*-2 (5'-CCGGGTGATCCAGCAC; SEQ ID NO:25). These oligos were annealed to one another and were cloned into the *Xma*I site of pMS2-2 (SenGupta et al., *Proc. Natl. Acad. Sci. USA* 93:8496, 1996). A number of clones were sequenced; one clone that had insert in the proper orientation was tested for transcriptional activation capability in yeast.

Results

In designing our riboactivator screening system, we first considered what length RNA we should screen. In order to answer this question, we considered what is known about RNA structure and protein-RNA interactions. One experimental system that has provided a significant amount of information about the probability that random RNAs will assume structures, and also about certain characteristics of protein-RNA interactions, is the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) system developed by Larry Gold and colleagues (reviewed in Gold et al., *Annu. Rev. Biochem.* 64:763, 1995). In this system, libraries of random nucleic acid molecules are synthesized *in vitro* and are screened, also *in vitro*, for their ability to interact with high affinity with particular targets. A sufficient number of SELEX experiments has been described to support the generalization that RNA transcripts with random base compositions are likely to assume some structure is they are at least 25 nucleotides long. These experiments have also shown that proteins that bind to 25 nt-long RNAs tend to interact with about 10-15 nts of the RNA. Although nothing

in these experiments indicates that RNAs *must* be 10-25 nt long to have secondary structure, or to have protein binding capability, we decided to utilize RNAs within this size range in our original screen.

In one particular SELEX experiment, researchers produced a library of RNA stem-loops and screened it to identify those that bind to an antibody at the antigen recognition site (Tsai et al., *Proc. Natl. Acad. Sci. USA* 89:8864, 1992). The RNA library was produced by inserting a random RNA 10-mer between two complementary sequences that could form a stable base-paired stem. Thus, all of the RNA molecules in the library contained the same stem. In some cases, the inserted 10-mer probably included additional self-complementary nucleotides, so that the stem was further extended; in others, the 10-mer probably remained as a loop.

Although it was clearly not our only option, we chose to utilize a stem-10mer-loop strategy analogous to that described above in our initial screen for riboactivators. We prepared a library of RNA molecules, termed the "R10 library" that each comprised two 15-nt complementary sequences flanking a random 10-mer RNA. Use of this library was attractive for two reasons. First, because the random RNA in the library is relatively short in sequence, it is feasible to screen every possible molecule. Also, it has been reported that 10 nt is the maximum size for an RNA loop *in vivo* (Wyatt et al. in *The RNA World* (Gestekand et al., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 465-496, 1993). Thus, when expressed *in vivo*, this library is likely to contain molecules with a variety of different stem-loop structures, ranging from structures with the minimum allowable loop size (thought to be 2 nt; Wyatt et al. in *The RNA World* (Gestekand et al., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 465-496, 1993) to the maximum

allowable loop size (10 nt). Presumably, bulged nucleotides are also not uncommon in the R10 library. Given that many RNA binding proteins recognize portions of RNA molecules that include loops or bulges (Draper, *Annu. Rev. Biochem.* 64:593, 1995), we expected that many molecules in the R10 population would be capable of interacting with proteins.

We transformed the R10 library into yeast, relying on the gap-repair method to integrate the library into a co-transformed plasmid containing the MS2 RNA so that, hybrid RNAs containing potential riboactivators linked to the MS2 RNA were produced *in vivo*. The yeast strain into which we transformed the library contained a gene for the lexA-MS2 fusion protein integrated in its genome, and also contained two different reporter genes, *HIS3* and *LacZ*, each of which is linked to a LexA operator. Transformants that contained riboactivators were selected for their ability to grow on a synthetic medium lacking histidine (i.e., for their ability to express *HIS3*). Aminotriazole, a competitive inhibitor of the *HIS3* gene product, was added to the medium to increase the stringency of the selection. Of approximately 8 million transformants, only approximately 200 expressed sufficient levels of *HIS3* to survive on the selective medium.

These 200 His+ transformants were screened to identify those that also expressed elevated levels of β-galactosidase; 12 were chosen as candidates after analysis by plate assay. Plasmids were isolated from these 12, were re-transformed into the same strain, and were re-assayed to ensure that the observed transcriptional activity was linked to the plasmid and did not reflect spurious mutation of the strain. 8 of these re-transformants proved to be plasmid-linked. We quantified the levels of activation that we were detecting by performing liquid β-galactosidase assays. We also isolated the plasmids from these 8 colonies and determined the nucleotide sequence of the RNAs they encoded. The β-galactosidase values and loop

sequences of our 8 positive clones are presented below in Table 1. As can be seen, there was a strong consensus of 5'-UGC(G>U>A)GG(U>A>C)(U>ACG)(C>A)(G>A>U)-3' (SEQ ID NO:4) among the activating RNAs.

TABLE 1

RIBOACTIVATOR ISOLATE NUMBER	SEQUENCE	β-GALACTOSIDASE
control (no insert)		0.2
1	5'-UGCGGGUACG SEQ ID NO:5	23.0
2	5'-UUGCUGGCGA SEQ ID NO:6	22.0
3	5'-UGCGGGUCAU SEQ ID NO:7	40.0
4	5'-UGCGGGUUCG SEQ ID NO:8	32.0
5	5'-UGCGGGAUCA SEQ ID NO:9	56.0
6	5'-UGCAGGUUCG SEQ ID NO:10	70.0
7	5'-UGCUGGAUCA SEQ ID NO:11	87.0
8	5'-UUGCUGGCGA SEQ ID NO:12	33.0

In performing our sequencing studies, we found that several of our riboactivating RNAs had alterations in sequences outside of the engineered stem-loop structure. Specifically, riboactivators numbered 3, 5, 6, and 7 had a deletion of a sequence element that is normally repeated in the MS2 RNA. Riboactivator number 7 also had a single base substitution (C to T) 15 nucleotides downstream of the stem-10bp structure.

We also sequenced eight non-activating RNAs in order to be sure that our library was not biased and that our riboactivator consensus truly reflected transcriptional regulatory activity rather than some other aspect of the library. Certain exemplary non-activator loop sequences are presented below:

TABLE 2

CONTROL RNA ISOLATE NUMBER	SEQUENCE
3	5'-CACGGTAAGT SEQ ID NO:13
6	5'-CAAAGACAGG SEQ ID NO:14
8	5'-GGCTGGTGGT SEQ ID NO:15
10	5'-GTAGAGCGA SEQ ID NO:16

Having identified our consensus sequence, we noticed that two known RNAs, the HIV Tar RNA and the RNA described above that was identified as having the ability to bind to an antibody raised against a particular 13-amino-acid peptide (Tsai et al., *Proc. Natl. Acad. Sci. USA* 89:8864, 1992), have similar nucleotide sequences to our consensus. Each of these RNAs has a stem-loop structure. The Tar loop is similar to our consensus but its stem is unrelated; the stem of the other RNA, which we refer to as the "Keene RNA", is identical to ours but its loop is less similar to our consensus than is Tar's. We tested whether either HIV Tar or the Keene RNA act as a riboactivator in our system. We also tested whether our riboactivator number 5 would retain its activation capability when positioned atop the Tar stem.

We found that neither Tar nor the Keene RNA activated transcription in our system. We further found that riboactivator number 5 retained significant activation capability when positioned atop the Tar stem. Table 3, below, summarizes our results with Tar.

TABLE 3

RNA	STEM SEQUENCE	LOOP SEQUENCE	β-GAL
vector	--	--	0
number 5	G-C U-A	CGGGAU	60
Tar	C-G G-C	CUGGGA	1
number 5 on Tar stem	C-G G-C	CGGGAU	20

We also tested whether riboactivator number 7 would retain its activity when positioned atop a different stem. Specifically, we placed the number 7 loop at the top of a 6-basepair stem comprising a *Sma*I site. As shown below in Table 4, we found that riboactivator number 7 retained significant activation capability when presented at the top of this *Sma*I stem.

TABLE 4

RNA	β-GALACTOSIDASE
number 7	80
number 7 on *Sma*I stem	40
control	0.1

In order to confirm that our riboactivators were working as expected, we transformed them into a control strain lacking the lexA-MS2 hybrid and confirmed that, in the absence of this hybrid (which is required to recruit the riboactivators to the DNA), the *LacZ* gene was not activated.

We performed three different experiments to ensure that our RNA molecules were responsible for the activation that we were observing. First, we re-cloned the 40 basepair segment encoding activators 7 and 5 into a fresh vector (containing both copies of the MS2 repeat and also lacking the base substitution that was present in original activator number 7) and confirming that the re-cloned activators retained transcriptional regulatory activity. Next, we re-cloned activators 5 and 7 in the "wrong" orientation and established that such "reverse" activators did not stimulate transcription. Finally, we made sure that our riboactivators could stimulate transcription when recruited to DNA by a DNA binding moiety other than the lexA DNA binding domain.

Specifically, we introduced the plasmids encoding our riboactivator/MS2 binding domain RNAs into a strain expressing a Ga14-MS2 hybrid and containing a reporter gene in which 5 Ga14 operators were positioned upstream of the *LacZ* gene. In this experiment, as shown below in Table 5, we found that our riboactivators gave the following levels of β-galactosidase expression:

TABLE 5

RIBOACTIVATOR ISOLATE NUMBER	β-GALACTOSIDASE
control (no insert)	20
1	43
2	38
3	37
4	42
5	58
6	58
7	75
8	46

In order to test whether our riboactivators functioned as "acidic activators", we re-transformed isolate number 7 into the L40 coat strain in combination with either a control vector (pFL46L) or a vector over-expressing Gal4 (pSS272-3). We tested the expression of the *LacZ* gene in the restraint strains by liquid β-galactosidase assay and found the following:

TABLE 6	
STRAIN	β-GALACTOSIDASE
control	30
Gal4 overexpression	9

As can be seen, our riboactivator was squelched by Gal4.

We also wanted to determine whether our riboactivators had sequences that were present elsewhere in the yeast genome. Since the entire yeast genome has been sequenced, this question can be answered definitively with a simple computer database search. We searched with both the exact sequence and its reverse complement for riboactivator number 7. The exact sequence was present in 12 locations, one of which was the *SWI1* gene, which encodes a component of the SWI/SNF global transcription activator complex. The reverse complement was present at 14 locations. Other riboactivator sequences were also found in the yeast genome, as summarized below in Table 7.

TABLE 7				
ACTIVATOR	GENE NAME	ACCESSION NO.	GENE FUNCTION/ STRUCTURE	ACTIVATOR SEQUENCE LOCATION
1	ALD7	Z75282	aldehyde dehydrogenase	151 nt from end of ORF

1	SEN3	Z37993	26s proteosome regulatory subunit	1685 nt from beginning of ORF; 1149 from end
2	YML093w	Z46660	similar to *P. falciparum* liver stage antigen LSA-1	421 nt from end of ORF
2	YMR006c	Z49810	hypothetical protein contains prenyl group binding (CAAX box)	1477 nt from end of ORF
2	YOR144c	Z75052	weak similarity to human DNA binding protein PO-GA	1070 nt from start of ORF
3	YOR064	Z74972	weak similarity to human retinoblastoma binding protein 2	89 nt past ORF
3	YBR044c	Z35913	similar to chaperonin HSP60 proteins	634 nt from start of ORF
4	PIE2	Z75017	phosphotidyl-inositol phosphate phosphatase; transmembrane	1489 nt from end of ORF
4	YOR306c	Z75214	similar to human X-linked pest-containing transporter	634 nt from start of ORF
5	ECM2	Z35934	involved in cell wall and plasma membrane biogenesis	268 nt from beginning of ORF

5	SCY1	Z72605	suppressor of GTPase mutant	189 nt from end of ORF
5	YDL057w	Z74105	hypothetical protein	58 nt after end of ORF
6	YNL187w	Z71463	hypothetical protein	14 nt from end of ORF
6	YPR105c	U32445	hypothetical transmembrane-spanning protein	444 nt from end of ORF
6	YAT1	X74553	carnitine acetyl transferase; mitochondrial transmembrane protein	363 nt from start of ORF
7	YLL013c	Z73118	similar to *Drosophila* pumilio protein	137 nt after beginning of ORF
7	TFC6	U28372	transcription factor; IIIc subunit Tau 91	345 nt after beginning of PRF
7	YLR376c	U19103	hypothetical transmembrane-spanning protein	526 nt past beginning of ORF
7	YBR203w	Z36072	hypothetical protein	1310 nt past ORF start
7	YGR257c	X99228	hypothetical protein with mitochondrial energy transfer protein signature	68 nt past ORF start
7	MIP1	Z28134	mitochondrial intermediate peptidase; transmembrane-spanning sequence	300 nt from end of ORF

7	YNL033w	Z71309	hypothetical transmembrane-spanning protein	615 nt past beginning of ORF
7	NUM1	Z50046	nuclear migration protein	6000 nt from start and 2000 nt from end of ORF
7	YOL092w	Z74834	protein similar to YBR147; transmembrane-spanning sequence	69 nt from start of ORF
7	SWI1	U33335	component of SWI/SNF global transcriptional activator complex	1600 nt from start of ORF
7	RFI2	Z49919	chromatin assembly complex, subunit p90	296 nt before start of ORF
7	YPR121w	U40828	similar to *B. subtilis* transcriptional activator tenA	152 nt past beginning of ORF

EXAMPLE 3

Identification of Riboactivators from "R40" library

Results

We wanted to prepare an RNA library that was not constrained, as is the above-described R10 library, by having all of its members contain the same base stem. Also, we were interested in preparing a library with a longer region of variability. Szostak and Ellington have recommended the use of RNAs 30-60 nts in length for formation of complex

structures such as multiple stem-loops, bulges, pseudoknots, and combinations thereof (Szostak et al. In *The RNA World* (Gesteland et al., eds), Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY, pp. 511-33. We therefore designed an oligonucleotide containing 40 randomized nucleotides flanked by the same sequences that flanked the stem-loop structure in the R10 library, so that the same primers could be used to amplify the R40 library as were used to amplify the R10 library. The R40 oligonucleotide has the sequence 5'-

CTCTGGGAGCTGCGATTGGCAGAATTCCGGCTAGAACTAGTGGATCCCCCNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGGCAGCTTGCATGCCTG
CAGGTCGACTCTAGAAAACATGAGGATCACCC (SEQ ID NO:17). The R40 library will then be transformed into yeast and screened as was the R10 library.

EXAMPLE 4

Identification of a Riboregulator Interaction Target

Materials and Methods

GEL MOBILITY SHIFT ASSAYS: Yeast nuclear extract was prepared according to standard procedures and was stored in Buffer A (25 mM Tris pH 7.6, 5 mM $MgCl_2$, 0.1 mM EDTA pH 8.0, 80 mM KCl, 10% gycerol, 1.0 mM DTT, 0.48M $(NH_4)_2SO_4$, 1mM ZnCl). Radiolabeled riboactivator number 7 and control RNA number 10 were prepared in the same buffer. Each RNA was incubated with yeast nuclear extract for 10 minutes at room temperature and the mixtures were separated on an acrylamide gel under non-denaturing conditions.

NORTHWESTERN ASSAYS: Northwestern blotting was performed according to Kwon et al. (*Dev. Biol.* 158:90, 1993). Briefly, 5-60 μg of protein extract was separated by SDS/PAGE and transferred to PVDF membranes (Biorad). Proteins were allowed to renature and then membranes were probed with radiolabeled number 7 riboactivator loop sequence (CCCGGGUGCUGGAUCACCCGGG; SEQ ID NO:26). After extensive washings, the filters were exposed to X-ray film so that protein-RNA interactions were detected.

Results

Our gel mobility shift assays indicated that riboactivator number 7, but not control RNA number 10, participated in a high molecular weight complex when incubated in yeast nuclear extract. Some complexes still formed when denatured extract was employed.

Northwestern analysis demonstrated that riboactivator number 7 bound to a protein in yeast nuclear extracts that was the size of yeast TBP (Figure 3). Further analyses with purified proteins revealed that number 7 interacts with yeast TBP, but not with yeast TFIID lacking TBP, with yeast RNA polymerase II holoenzyme, or with mammalian TFIID (Figure 4). The Keene RNA, which does not activate transcription (see Example 2 above), also appears to bind to a protein in yeast nuclear extract that is the same size as TBP.

Other Embodiments

Those of ordinary sill in the art will recognize that the foregoing constitutes a description merely of certain preferred embodiments of the invention and is not intended to define or limit the scope of the following claims. Various changes, substitutions,

modifications, and extensions can readily be achieved without undue experimentation and are intended to be within the scope of the claims.